

Weighted graphs with distances in given ranges

Elena Rubei

Abstract

Let $\mathcal{G} = (G, w)$ be a weighted simple finite connected graph, that is, let G be a simple finite connected graph endowed with a function w from the set of the edges of G to the set of real numbers. For any subgraph G' of G , we define $w(G')$ to be the sum of the weights of the edges of G' . For any i, j vertices of G , we define $D_{\{i,j\}}(\mathcal{G})$ to be the minimum of the weights of the simple paths of G joining i and j . The $D_{\{i,j\}}(\mathcal{G})$ are called 2-weights of \mathcal{G} . Weighted graphs and their reconstruction from 2-weights have applications in several disciplines, such as biology and psychology.

Let $\{m_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ and $\{M_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ be two families of positive real numbers parametrized by the 2-subsets of $\{1, \dots, n\}$ with $m_I \leq M_I$ for any I ; we study when there exist a positive-weighted graph \mathcal{G} and an n -subset $\{1, \dots, n\}$ of the set of its vertices such that $D_I(\mathcal{G}) \in [m_I, M_I]$ for any $I \in \binom{\{1, \dots, n\}}{2}$. Then we study the analogous problem for trees, both in the case of positive weights and in the case of general weights.

1 Introduction

For any graph G , let $E(G)$, $V(G)$ and $L(G)$ be respectively the set of the edges, the set of the vertices and the set of the leaves of G . A **weighted graph** $\mathcal{G} = (G, w)$ is a graph G endowed with a function $w : E(G) \rightarrow \mathbb{R}$. For any edge e , the real number $w(e)$ is called the weight of the edge. If all the weights are nonnegative (respectively positive), we say that the graph is **nonnegative-weighted** (respectively **positive-weighted**). Throughout the paper we will consider only simple finite connected graphs.

For any subgraph G' of G , we define $w(G')$ to be the sum of the weights of the edges of G' .

Definition 1. Let $\mathcal{G} = (G, w)$ be a weighted graph. For any distinct $i, j \in V(G)$, we define

$$D_{\{i,j\}}(\mathcal{G}) = \min\{w(p) \mid p \text{ a simple path of } G \text{ joining } i \text{ and } j\}.$$

More simply, we denote $D_{\{i,j\}}(\mathcal{G})$ by $D_{i,j}(\mathcal{G})$ for any order of i, j . We call the $D_{i,j}(\mathcal{G})$ the **2-weights** (or **distances**) of \mathcal{G} .

Observe that in the case \mathcal{G} is a tree, $D_{i,j}(\mathcal{G})$ is the weight of the unique path joining i and j .

If S is a subset of $V(G)$, the 2-weights give a vector in $\mathbb{R}^{\binom{S}{2}}$. This vector is called **2-dissimilarity vector** of (\mathcal{G}, S) . Equivalently, we can speak of the **family of the 2-weights** of (\mathcal{G}, S) .

We can wonder when a family of real numbers is the family of the 2-weights of some weighted graph and of some subset of the set of its vertices. If S is a finite set of cardinality greater than 2, we say that a family

2010 Mathematical Subject Classification: 05C05, 05C12, 05C22

Key words: weighted graphs, distances, range

of real numbers $\{D_I\}_{I \in \binom{S}{2}}$ is **graphlike** (respectively p-graphlike, nn-graphlike) if there exist a weighted graph (respectively a positive-weighted graph, a nonnegative-weighted graph) $\mathcal{G} = (G, w)$ and a subset S of the set of its vertices such that $D_I(\mathcal{G}) = D_I$ for any 2-subset I of S . If the graph is a weighted (respectively positive-weighted, nonnegative-weighted) tree $\mathcal{T} = (T, w)$ we say that the family is **treelike** (respectively p-treelike, nn-treelike). If, in addition, $S \subset L(T)$, we say that the family is **l-treelike** (respectively, p-l-treelike, nn-l-treelike).

Weighted graphs have applications in several disciplines, such as biology and psychology. Phylogenetic trees are weighted graphs whose vertices represent species and the weight of an edge is given by how much the DNA sequences of the species represented by the vertices of the edge differ. Dissimilarity families arise naturally also in psychology, see for instance the introduction in [8]. There is a wide literature concerning graphlike dissimilarity families and treelike dissimilarity families, in particular concerning methods to reconstruct weighted trees from their dissimilarity families; these methods are used by biologists to reconstruct phylogenetic trees. See for example [15], [21] and [9], [19] for overviews.

The first contribution to the characterization of graphlike families of numbers dates back to 1965 and it is due to Hakimi and Yau, see [12]:

Theorem 2. (Hakimi-Yau) *A family of positive real numbers $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ is p-graphlike if and only if the D_I satisfy the triangle inequalities, i.e. if and only if $D_{i,j} \leq D_{i,k} + D_{k,j}$ for any distinct $i, j, k \in [n]$.*

In the same years, also a criterion for a metric on a finite set to be nn-l-treelike was established, see [6], [20], [22]:

Theorem 3. (Buneman-SimoesPereira-Zaretskii) *Let $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ be a set of positive real numbers satisfying the triangle inequalities. It is p-treelike (or nn-l-treelike) if and only if, for all distinct $i, j, k, h \in \{1, \dots, n\}$, the maximum of*

$$\{D_{i,j} + D_{k,h}, D_{i,k} + D_{j,h}, D_{i,h} + D_{k,j}\}$$

is attained at least twice.

Also the case of not necessarily nonnegative weights has been studied. In 1972 Hakimi and Patrinos proved the following theorem (see [11]):

Theorem 4. (Hakimi-Patrinos) *A family of real numbers $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ is always the family of the 2-weights of some weighted graph and some subset $\{1, \dots, n\}$ of its vertices*

In [4], Bandelt and Steel proved a result, analogous to Theorem 3, for general weighted trees:

Theorem 5. (Bandelt-Steel) *For any set of real numbers $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$, there exists a weighted tree \mathcal{T} with leaves $1, \dots, n$ such that $D_I(\mathcal{T}) = D_I$ for any 2-subset I of $\{1, \dots, n\}$ if and only if, for any distinct $a, b, c, d \in \{1, \dots, n\}$, we have that at least two among*

$$D_{a,b} + D_{c,d}, \quad D_{a,c} + D_{b,d}, \quad D_{a,d} + D_{b,c}$$

are equal.

Recently Baldisserrri characterized the families $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ that are the families of the 2-weights of positive-weighted trees with exactly n vertices, see [1].

Finally we want to mention that recently k -weights of weighted graphs for $k \geq 3$ have been introduced and studied; in particular there are some results concerning the characterization of families of k -weights, see for instance [5], [2], [3], [10], [13], [14], [16], [17], and [18].

In this paper, we study when there exists a weighted graph with 2-weights in given ranges; this problem can be of interest because the data one can get from experiments are obviously not precise, on the contrary they can vary in a range. Precisely, let $\{m_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ and $\{M_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ be two families of positive real numbers parametrized by the 2-subsets of $\{1, \dots, n\}$ with $m_I \leq M_I$ for any I ; in §3 we study when there exist a weighted graph \mathcal{G} and an n -subset $\{1, \dots, n\}$ of the set of its vertices such that $D_I(\mathcal{G}) \in [m_I, M_I]$ for any $I \in \binom{\{1, \dots, n\}}{2}$. Finally, in §4 we study the analogous problem for trees, both in the case of positive weights and in the case of general weights. The treatment of the case of trees turns out to be much more complicated and long than the case of graphs.

2 Preliminaries

Notation 6. • For any $n \in \mathbb{N} - \{0\}$, let $[n] = \{1, \dots, n\}$.

- For any set S and $k \in \mathbb{N}$, let $\binom{S}{k}$ be the set of the k -subsets of S .
- For any family of real numbers or unknowns parametrized by $\binom{[n]}{2}$, $\{x_{\{i,j\}}\}_{\{i,j\} \in \binom{[n]}{2}}$, we denote $x_{\{i,j\}}$ by $x_{i,j}$ for any order of i and j .
- Throughout the paper, the word “graph” will denote a finite simple connected graph.
- Let T be a tree and let S be a subset of $L(T)$. We denote by $T|_S$ the minimal subtree of T whose set of vertices contains S .
- Let T be a tree. We say that two leaves i and j of T are neighbours if in the path joining i and j there is only one vertex of degree greater than or equal to 3.

The following theorem (see [7]) and the following lemma will be useful to solve our problem in the case of trees.

Theorem 7. (Carver) Let $L_i(x_1, \dots, x_t)$ for $i = 1, \dots, s$ be polynomials of degree 1 in x_1, \dots, x_t . The system of inequalities

$$\begin{cases} L_1(x_1, \dots, x_t) > 0 \\ \dots \\ L_s(x_1, \dots, x_t) > 0 \end{cases}$$

is solvable if and only if there does not exist a set of $s + 1$ constants, c_1, \dots, c_{s+1} , such that

$$\sum_{i=1, \dots, s} c_i L_i(x_1, \dots, x_t) + c_{s+1} \equiv 0,$$

at least one of the c 's being positive and none of them being negative.

Remark 8. Let S be a system of linear inequalities in x_1, \dots, x_t . We can write it as follows:

$$\left\{ \begin{array}{l} x_t > L_1(x_1, \dots, x_{t-1}) \\ \dots\dots\dots \\ x_t > L_s(x_1, \dots, x_{t-1}) \\ x_t < M_1(x_1, \dots, x_{t-1}) \\ \dots\dots\dots \\ x_t < M_r(x_1, \dots, x_{t-1}) \\ N_1(x_1, \dots, x_{t-1}) > 0 \\ \dots\dots\dots \\ N_p(x_1, \dots, x_{t-1}) > 0 \end{array} \right.$$

for some linear polynomials L_i, M_j, N_l in x_1, \dots, x_{t-1} . The system S is solvable if and only if the system in x_1, \dots, x_{t-1} given by the inequalities

$$M_j(x_1, \dots, x_{t-1}) > L_i(x_1, \dots, x_{t-1}),$$

for $i = 1, \dots, s$, $j = 1, \dots, r$, and the inequalities

$$N_l(x_1, \dots, x_{t-1}) > 0,$$

for $l = 1, \dots, p$, is solvable. We get an analogous statement if we replace some of the strict inequalities with nonstrict inequalities.

Lemma 9. Let $z_1, \dots, z_s, t \in \mathbb{N} - \{0\}$ and let $L_i(x_1, \dots, x_t)$ for $i = 1, \dots, r$ be polynomials of degree 1 in the unknowns x_1, \dots, x_t . If, for any $\varepsilon > 0$, the system

$$\left\{ \begin{array}{l} L_1(x_1, \dots, x_t) > -z_1\varepsilon \\ \dots\dots\dots \\ L_s(x_1, \dots, x_t) > -z_s\varepsilon \\ L_{s+1}(x_1, \dots, x_t) \geq 0 \\ \dots\dots\dots \\ L_r(x_1, \dots, x_t) \geq 0 \end{array} \right. \quad (1)$$

is solvable, then also the system

$$\left\{ \begin{array}{l} L_1(x_1, \dots, x_t) \geq 0 \\ \dots\dots\dots \\ L_s(x_1, \dots, x_t) \geq 0 \\ L_{s+1}(x_1, \dots, x_t) \geq 0 \\ \dots\dots\dots \\ L_r(x_1, \dots, x_t) \geq 0 \end{array} \right. \quad (2)$$

is solvable.

Proof. We prove the statement by induction on t .

The statement in the case $t = 1$ is easy to prove. Let us prove the induction step $t - 1 \Rightarrow t$. Suppose that, for any $\varepsilon > 0$, the system (1) is solvable; then also the system (in x_1, \dots, x_{t-1}) we get from it by “eliminating” the unknown x_t (see Remark 8) is solvable. By induction assumption also the system we get from it by replacing $>$ with \geq and putting $\varepsilon = 0$ is solvable. But this last system is exactly the system we get from (2) by eliminating the unknown x_t . So also (2) is solvable. \square

3 The case of graphs

Theorem 10. *Let $\{m_I\}_{I \in \binom{[n]}{2}}$ and $\{M_I\}_{I \in \binom{[n]}{2}}$ be two families of positive real numbers with $m_I \leq M_I$ for any I . There exist a positive-weighted graph \mathcal{G} and an n -subset $[n]$ of the set of its vertices such that $D_I(\mathcal{G}) \in [m_I, M_I]$ for any $I \in \binom{[n]}{2}$ if and only if for any $i, j \in [n]$ with $i \neq j$ we have*

$$m_{i,j} \leq M_{i,t_1} + M_{t_1,t_2} + \dots + M_{t_{k-1},t_k} + M_{t_k,j}$$

for any $k \in \mathbb{N}$ and $t_1, \dots, t_k \in [n] - \{i, j\}$ with $t_\alpha \neq t_{\alpha+1}$ for any $\alpha = 1, \dots, k - 1$.

Proof. \Rightarrow Suppose there exist a positive-weighted graph \mathcal{G} and an n -subset $\{1, \dots, n\}$ of the set of its vertices such that $D_I(\mathcal{G}) \in [m_I, M_I]$ for any $I \in \binom{[n]}{2}$. We recall that, for the $D_I(\mathcal{G})$, the triangle inequalities hold, see Theorem 2. Then, for any $i, j \in [n]$ with $i \neq j$,

$$\begin{aligned} m_{i,j} \leq D_{i,j}(\mathcal{G}) &\leq D_{i,t_1}(\mathcal{G}) + D_{t_1,j}(\mathcal{G}) \leq \\ &\leq \dots \leq \\ &\leq D_{i,t_1}(\mathcal{G}) + D_{t_1,t_2}(\mathcal{G}) + \dots + D_{t_{k-1},t_k}(\mathcal{G}) + D_{t_k,j}(\mathcal{G}) \leq \\ &\leq M_{i,t_1} + M_{t_1,t_2} + \dots + M_{t_{k-1},t_k} + M_{t_k,j} \end{aligned}$$

for any $k \in \mathbb{N}$ and $t_1, \dots, t_k \in [n] - \{i, j\}$ with $t_\alpha \neq t_{\alpha+1}$ for any $\alpha = 1, \dots, k - 1$.

\Leftarrow Let us define, for any $i, j \in [n]$ with $i \neq j$,

$$\tilde{M}_{i,j} := \min\{M_{i,t_1} + M_{t_1,t_2} + \dots + M_{t_{k-1},t_k} + M_{t_k,j}\}_{k \in \mathbb{N}, t_1, \dots, t_k \in [n] - \{i, j\}, t_\alpha \neq t_{\alpha+1} \forall \alpha = 1, \dots, k-1}.$$

It is easy to see that the $\tilde{M}_{i,j}$ satisfy the triangle inequalities $\tilde{M}_{i,j} \leq \tilde{M}_{i,o} + \tilde{M}_{o,j}$ for any distinct $i, j, o \in [n]$, in fact:

$$\begin{aligned} \min\{M_{i,t_1} + M_{t_1,t_2} + \dots + M_{t_{k-1},t_k} + M_{t_k,j}\}_{k \in \mathbb{N}, t_1, \dots, t_k \in [n] - \{i, j\}, t_\alpha \neq t_{\alpha+1} \forall \alpha = 1, \dots, k-1} \leq \\ \leq M_{i,v_1} + M_{v_1,v_2} + \dots + M_{v_{r-1},v_r} + M_{v_r,o} + M_{o,w_1} + M_{w_1,w_2} + \dots + M_{w_{s-1},w_s} + M_{w_s,j} \end{aligned}$$

for any $r, s \in \mathbb{N}$, $v_\alpha \in [n] - \{i, o\}$ for $\alpha = 1, \dots, r$, $v_\alpha \neq v_{\alpha+1}$ for $\alpha = 1, \dots, r - 1$, $w_\alpha \in [n] - \{o, j\}$ for $\alpha = 1, \dots, s$, $w_\alpha \neq w_{\alpha+1}$ for $\alpha = 1, \dots, s - 1$ (consider two cases: the case where no one of the v_α and the w_α is in $\{i, j\}$ and the case where at least one of the v_α or the w_α is in $\{i, j\}$). So, by Theorem 2, there exists a positive-weighted graph \mathcal{G} such that $D_{i,j}(\mathcal{G}) = \tilde{M}_{i,j}$ for any $i, j \in [n]$ with $i \neq j$. By our assumption, we have that $\tilde{M}_{i,j} \geq m_{i,j}$ for any $i, j \in [n]$ with $i \neq j$ and obviously $\tilde{M}_{i,j} \leq M_{i,j}$ for any $i, j \in [n]$ with $i \neq j$, so we conclude. \square

4 The case of trees

Definition 11. Let X be a set and let Y be a 4-subset of X (a quartet). A **(quartet) split** of Y is a partition of Y into two disjoint 2-subsets. We denote the split $\{\{a, b\}, \{c, d\}\}$ simply by $(a, b \mid c, d)$.

Let S be a system (that is, a set) of splits of the quartets of X .

We say that S is **fat** if, for every quartet of X , either exactly one of its splits or all its splits are in S .

Following [9], Ch. 3, we say that S is **transitive** if, for any distinct $a, b, c, d, e \in X$, the following implication holds:

$$(a, b \mid c, d) \in S \text{ and } (a, b \mid c, e) \in S \implies (a, b \mid d, e) \in S.$$

Following again [9], we say that S is **saturated** if, for any distinct $a_1, a_2, b_1, b_2, x \in X$, the following implication holds:

$$(a_1, a_2 \mid b_1, b_2) \in S \implies \text{either } (a_1, x \mid b_1, b_2) \in S \text{ or } (a_1, a_2 \mid b_1, x) \in S.$$

The statement of the following lemma is similar to the characterization of the system of the splits of the quartets coming from trees (with a slight difference in the definition of the splits of a quartet of leaves of a tree), see [9] Thm. 3.7 and [8].

Lemma 12. Let $n \in \mathbb{N}$, $n \geq 4$. Let S be a system of splits of the quartets of $[n]$. Suppose S is fat, transitive and saturated. Then the linear system in the unknowns x_I for $I \in \binom{[n]}{2}$ given by the equations

$$x_{a,c} - x_{b,c} = x_{a,d} - x_{b,d}$$

for any $(a, b \mid c, d) \in S$ has a nonzero solution.

Proof. We prove the statement by induction on n . If $n = 4$, the statement is obvious. Let us prove the induction step. Suppose that D_I for $I \in \binom{[n-1]}{2}$ solve the equations

$$x_{a,c} - x_{b,c} = x_{a,d} - x_{b,d}$$

for any $(a, b \mid c, d) \in S$ with $a, b, c, d \in [n-1]$ and that they are not all zero. We want to find $D_{n,i}$ for $i = 1, \dots, n-1$ such that the D_I for $I \in \binom{[n]}{2}$ solve the linear system given by all the elements of S .

Let us define $D_{n,1}$ at random.

Let us define $D_{n,2}$ as follows:

if there does not exist $x \in [n-1] - \{1, 2\}$ such that $(n, x \mid 1, 2) \in S$, we define $D_{n,2}$ at random;

if there exists $x \in [n-1] - \{1, 2\}$ such that $(n, x \mid 1, 2) \in S$, we set

$$D_{n,2} := D_{n,1} + D_{x,2} - D_{x,1};$$

it is a good definition, in fact if there exists $y \in [n-1] - \{x, 1, 2\}$ such that $(n, y \mid 1, 2) \in S$, we have that, by the transitivity of S , $(x, y \mid 1, 2) \in S$, so

$$D_{n,1} + D_{x,2} - D_{x,1} = D_{n,1} + D_{y,2} - D_{y,1}.$$

In an analogous way we define the other $D_{n,i}$; precisely, suppose we have defined $D_{n,1}, \dots, D_{n,k-1}$ in such a way that $D_{n,1}, \dots, D_{n,k-1}$ and $D_{i,j}$ for $i, j \in [n-1]$ satisfy the equations induced by S involving $x_{n,1}, \dots, x_{n,k-1}$ and $x_{i,j}$ for $i, j \in [n-1]$; we define $D_{n,k}$ as follows:

if there do not exist $x \in [n-1]$ and $i \in [k-1]$ with $x \neq k, i$ and such that $(n, x \mid k, i) \in S$, we define $D_{n,k}$ at random;

if there exist $x \in [n-1]$ and $i \in [k-1]$ with $x \neq k, i$ such that $(n, x \mid k, i) \in S$, we set

$$D_{n,k} := D_{n,i} + D_{x,k} - D_{x,i}.$$

We have to show that it is a good definition. Suppose $y \in [n-1]$ and $j \in [k-1]$ with $y \neq k, j$ are such that $(n, y \mid k, j) \in S$; we have to show that

$$D_{n,i} + D_{x,k} - D_{x,i} = D_{n,j} + D_{y,k} - D_{y,j}. \quad (3)$$

Since S is saturated and transitive, from $(n, x \mid k, i) \in S$, we get:

either

$$(n, y \mid k, i) \in S \quad \text{and} \quad (x, y \mid k, i) \in S \quad (4)$$

or

$$(n, x \mid k, y) \in S \quad \text{and} \quad (n, x \mid y, i) \in S. \quad (5)$$

From $(n, x \mid k, i) \in S$, we get:

either

$$(n, j \mid k, i) \in S \quad \text{and} \quad (x, j \mid k, i) \in S \quad (6)$$

or

$$(n, x \mid k, j) \in S \quad \text{and} \quad (n, x \mid j, i) \in S. \quad (7)$$

From $(n, y \mid k, j) \in S$, we get:

either

$$(n, x \mid k, j) \in S \quad \text{and} \quad (x, y \mid k, j) \in S \quad (8)$$

or

$$(n, y \mid k, x) \in S \quad \text{and} \quad (n, y \mid x, j) \in S. \quad (9)$$

Finally, from $(n, y \mid k, j) \in S$, we get:

either

$$(n, i \mid k, j) \in S \quad \text{and} \quad (y, i \mid k, j) \in S \quad (10)$$

or

$$(n, y \mid k, i) \in S \quad \text{and} \quad (n, y \mid i, j) \in S. \quad (11)$$

If condition (8) holds, we get, from it and from the assumption $(n, x \mid k, i) \in S$, that also $(n, x \mid i, j) \in S$ holds (by the transitivity of S). So the statement (3) is equivalent to the equality

$$D_{x,i} + D_{x,k} - D_{x,i} = D_{x,j} + D_{y,k} - D_{y,j},$$

which follows from $(x, y \mid k, j) \in S$.

If condition (4) holds, we get our statement in an analogous way (swap i with j and x with y).

If condition (11) holds, we get, from it and from the assumption $(n, x \mid k, i) \in S$, that also $(x, y \mid k, i) \in S$ holds. From the condition that $(n, y \mid i, j) \in S$, the statement (3) is equivalent to the equality

$$D_{y,i} + D_{x,k} - D_{x,i} = D_{y,j} + D_{y,k} - D_{y,j},$$

which follows from $(x, y \mid k, i) \in S$.

If condition (7) holds, we get our statement in an analogous way (swap i with j and x with y).

So we can suppose that (9), (5), (10), (6) hold. From the fact that $(n, j \mid k, i) \in S$ (which is true by (6)), the fact that $(n, i \mid k, j) \in S$ (which is true by (10)) and the fatness of S , we get that $(n, k \mid i, j) \in S$. From the condition that $(x, j \mid k, i) \in S$ (which is true by (6)), the statement (3) is equivalent to the equality

$$D_{n,i} + D_{j,k} - D_{j,i} = D_{n,j} + D_{y,k} - D_{y,j}.$$

By the condition $(n, k \mid i, j) \in S$, this equality is equivalent to

$$D_{k,i} + D_{j,k} - D_{j,i} = D_{k,j} + D_{y,k} - D_{y,j}.$$

which is true since $(i, y \mid k, j) \in S$ (which follows from (10)). \square

Theorem 13. *Let $\{m_I\}_{I \in \binom{[n]}{2}}$ and $\{M_I\}_{I \in \binom{[n]}{2}}$ be two families of real numbers with $m_I < M_I$ for any I . There exists a weighted tree $\mathcal{T} = (T, w)$ with $L(T) = [n]$ and such that $D_I(\mathcal{T}) \in (m_I, M_I)$ for any $I \in \binom{[n]}{2}$ if and only if there exists a set S of splits of the quartets of $[n]$ such that*

- (i) S is fat, transitive and saturated,
- (ii)

$$m_{\sigma_1} + \dots + m_{\sigma_r} < M_{\tau_1} + \dots + M_{\tau_r}$$

for any $r \in \mathbb{N} - \{0\}$, for any $(\sigma_1, \dots, \sigma_r)$ and (τ_1, \dots, τ_r) partitions of the same $2r$ -subset of $[n]$ into 2-sets such that $(\sigma_1, \dots, \sigma_r)$ can be obtained from (τ_1, \dots, τ_r) with transformations on the 2-sets of the following kind:

$$(i, k), (j, l) \mapsto (i, j), (k, l)$$

for any $(j, k \mid i, l) \in S$.

Proof. \Rightarrow Let $\mathcal{T} = (T, w)$ be a weighted tree with $L(T) = [n]$ and such that $D_I(\mathcal{T}) \in (m_I, M_I)$ for any $I \in \binom{[n]}{2}$. We define S in the following way: for any quartet $\{a, b, c, d\}$ in $[n]$, we say that $(a, b \mid c, d) \in S$ if and only if a and b are neighbours and c and d are neighbours in $T|_{a,b,c,d}$. It is easy to see that S is fat, transitive and saturated. Furthermore, for any $(\sigma_1, \dots, \sigma_r)$ and (τ_1, \dots, τ_r) partitions of the same subset of $[n]$ into 2-sets such that $(\sigma_1, \dots, \sigma_r)$ can be obtained from (τ_1, \dots, τ_r) with transformations on the 2-sets of the kind $(i, k \mid j, l) \mapsto (i, j \mid k, l)$ for any $(j, k \mid i, l) \in S$, we have:

$$m_{\sigma_1} + \dots + m_{\sigma_r} < D_{\sigma_1}(\mathcal{T}) + \dots + D_{\sigma_r}(\mathcal{T}) = D_{\tau_1}(\mathcal{T}) + \dots + D_{\tau_r}(\mathcal{T}) < M_{\tau_1} + \dots + M_{\tau_r},$$

hence (ii) holds.

\Leftarrow By Lemma 12, the linear system given by the equations

$$D_{a,c} - D_{b,c} = D_{a,d} - D_{b,d}$$

for any $(a, b \mid c, d) \in S$ has nonzero solutions. So we can write some unknowns, D_{I_1}, \dots, D_{I_s} , in function of some others: D_{J_1}, \dots, D_{J_t} for some $t \geq 1$: let

$$D_{I_i} = f_{I_i}(D_{J_1}, \dots, D_{J_t})$$

for $i = 1, \dots, s$. Consider the following system of inequalities in D_{J_1}, \dots, D_{J_t} :

$$\left\{ \begin{array}{l} D_{J_1} - m_{J_1} > 0 \\ \dots\dots\dots \\ D_{J_t} - m_{J_t} > 0 \\ f_{I_1}(D_{J_1}, \dots, D_{J_t}) - m_{I_1} > 0 \\ \dots\dots\dots \\ f_{I_s}(D_{J_1}, \dots, D_{J_t}) - m_{I_s} > 0 \\ -D_{J_1} + M_{J_1} > 0 \\ \dots\dots\dots \\ -D_{J_t} + M_{J_t} > 0 \\ -f_{I_1}(D_{J_1}, \dots, D_{J_t}) + M_{I_1} > 0 \\ \dots\dots\dots \\ -f_{I_s}(D_{J_1}, \dots, D_{J_t}) + M_{I_s} > 0 \end{array} \right. \quad (12)$$

By condition (ii) there does not exist a set of $2t + 2s + 1$ nonnegative constants, $c_1, \dots, c_{2t+2s+1}$, with at least one of them positive, such that the linear combination of the first members of the inequalities of (12) with coefficients c_1, \dots, c_{2t+2s} plus $c_{2t+2s+1}$ is identically zero. So, by Carver's Theorem, the system (12) is solvable. Let $(\overline{D}_{J_1}, \dots, \overline{D}_{J_t})$ be a solution. By the fatness of S and by Theorem 5, for the dissimilarity vector with entries

$$\overline{D}_{J_1}, \dots, \overline{D}_{J_t}, \overline{D}_{I_1} := f_{I_1}(\overline{D}_{J_1}, \dots, \overline{D}_{J_t}), \dots, \overline{D}_{I_s} := f_{I_s}(\overline{D}_{J_1}, \dots, \overline{D}_{J_t}),$$

there exists a weighted tree $\mathcal{T} = (T, w)$ with $L(T) = [n]$ and such that $D_I(\mathcal{T}) = \overline{D}_I$ for any $I \in \binom{\{1, \dots, n\}}{2}$, so $D_I(\mathcal{T}) \in (m_I, M_I)$ for any $I \in \binom{\{1, \dots, n\}}{2}$. \square

Remark 14. Observe that the same technique can be useful to study the analogous problem for some kind of tree. For instance we can prove easily in an analogous way that, given two families of real numbers, $\{m_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ and $\{M_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ with $m_I < M_I$ for any I , there exists a weighted star $\mathcal{T} = (T, w)$ with $L(T) = [n]$ and such that $D_I(\mathcal{T}) \in (m_I, M_I)$ for any $I \in \binom{\{1, \dots, n\}}{2}$ if and only if

$$m_{\sigma_1} + \dots + m_{\sigma_r} < M_{\tau_1} + \dots + M_{\tau_r}$$

for any $(\sigma_1, \dots, \sigma_r)$ and (τ_1, \dots, τ_r) partitions of the same subset of $[n]$ into 2-sets.

Considering 2-weights in closed intervals, we get the following theorem.

Theorem 15. Let $\{m_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ and $\{M_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ be two families of real numbers with $m_I \leq M_I$ for any I . There exists a weighted tree $\mathcal{T} = (T, w)$ with $L(T) = [n]$ and such that $D_I(\mathcal{T}) \in [m_I, M_I]$ for any $I \in \binom{\{1, \dots, n\}}{2}$ if and only if there exists a system S of splits of the quartets of $[n]$ such that

(i) S is fat, transitive and saturated,

(ii)

$$m_{\sigma_1} + \dots + m_{\sigma_r} \leq M_{\tau_1} + \dots + M_{\tau_r}$$

for any $(\sigma_1, \dots, \sigma_r)$ and (τ_1, \dots, τ_r) partitions of the same subset of $[n]$ into 2-sets such that $(\sigma_1, \dots, \sigma_r)$ can be obtained from (τ_1, \dots, τ_r) with transformations on the 2-sets of the following kind:

$$(i, k), (j, l) \mapsto (i, j), (k, l)$$

for any $(j, k \mid i, l) \in S$.

Proof. The proof of the implication \Rightarrow is completely analogous to the proof of the same implication of Theorem 13. Let us prove the other implication. By Lemma 12, the linear system given by the equations $D_{a,c} - D_{b,c} = D_{a,d} - D_{b,d}$ for any $(a, b \mid c, d) \in S$ has nonzero solutions. So we can write some unknowns, D_{I_1}, \dots, D_{I_r} , in function of some others D_{J_1}, \dots, D_{J_t} for some $t \geq 1$: let

$$D_{I_i} = f_{I_i}(D_{J_1}, \dots, D_{J_t})$$

for any $i = 1, \dots, r$. Consider the system of inequalities

$$\left\{ \begin{array}{l} D_{J_1} - m_{J_1} + \epsilon > 0 \\ \dots\dots\dots \\ D_{J_t} - m_{J_t} + \epsilon > 0 \\ f_{I_1}(D_{J_1}, \dots, D_{J_t}) - m_{I_1} + \epsilon > 0 \\ \dots\dots\dots \\ f_{I_s}(D_{J_1}, \dots, D_{J_t}) - m_{I_s} + \epsilon > 0 \\ -D_{J_1} + M_{J_1} + \epsilon > 0 \\ \dots\dots\dots \\ -D_{J_t} + M_{J_t} + \epsilon > 0 \\ -f_{I_1}(D_{J_1}, \dots, D_{J_t}) + M_{I_1} + \epsilon > 0 \\ \dots\dots\dots \\ -f_{I_s}(D_{J_1}, \dots, D_{J_t}) + M_{I_s} + \epsilon > 0 \end{array} \right. \quad (13)$$

By condition (ii), we have that, for any $\epsilon > 0$,

$$m_{\sigma_1} + \dots + m_{\sigma_r} - (2r)\epsilon < M_{\tau_1} + \dots + M_{\tau_r}$$

for any $r \in \mathbb{N} - \{0\}$, for any $(\sigma_1, \dots, \sigma_r)$ and (τ_1, \dots, τ_r) partitions of the same $2r$ -subset of $[n]$ into 2-sets such that $(\sigma_1, \dots, \sigma_r)$ can be obtained from (τ_1, \dots, τ_r) with transformations on the 2-sets of the kind $(i, k \mid j, l) \mapsto (i, j \mid k, l)$ for any $(j, k \mid i, l) \in S$. So there does not exist a set of $2t + 2s + 1$ nonnegative constants, $c_1, \dots, c_{2t+2s+1}$ with at least one of them positive, such that the linear combination of the first members of the inequalities of (13) with coefficients c_1, \dots, c_{2t+2s} plus $c_{2t+2s+1}$ is identically zero. So, by Carver's theorem, the system (13) is solvable for any $\epsilon > 0$. Hence, by Lemma 9, the system we get from (13) by replacing \geq with $>$ and ϵ with 0 is solvable. Let $(\overline{D}_{J_1}, \dots, \overline{D}_{J_t})$ be a solution.

By the fatness of S and by Theorem 5, for the dissimilarity vector with entries

$$\overline{D}_{J_1}, \dots, \overline{D}_{J_t}, \overline{D}_{I_1} := f_{I_1}(\overline{D}_{J_1}, \dots, \overline{D}_{J_t}), \dots, \overline{D}_{I_s} := f_{I_s}(\overline{D}_{J_1}, \dots, \overline{D}_{J_t}),$$

there exists a weighted tree $\mathcal{T} = (T, w)$ with $L(T) = [n]$ and such that $D_I(\mathcal{T}) = \overline{D}_I$ for any $I \in \binom{[n]}{2}$, so $D_I(\mathcal{T}) \in [m_I, M_I]$ for any $I \in \binom{[n]}{2}$. □

By using Theorem 3, we get a theorem, analogous to the previous ones, for positive-weighted trees:

Theorem 16. *Let $\{m_I\}_{I \in \binom{[n]}{2}}$ and $\{M_I\}_{I \in \binom{[n]}{2}}$ be two families of positive real numbers with $m_I < M_I$ for any I . There exists a positive-weighted tree $\mathcal{T} = (T, w)$ with $L(T) = [n]$ and such that $D_I(\mathcal{T}) \in (m_I, M_I)$ for any $I \in \binom{[n]}{2}$ if and only if there exists a system S of splits of the quartets of $[n]$ such that the condition (i) of Theorem 13 and the following condition hold:*

(ii)

$$m_{\sigma_1} + \dots + m_{\sigma_r} < M_{\tau_1} + \dots + M_{\tau_s}$$

for any $(\sigma_1, \dots, \sigma_r)$ and (τ_1, \dots, τ_s) partitions of subsets of $[n]$ into 2-sets such that (τ_1, \dots, τ_s) can be obtained from $(\sigma_1, \dots, \sigma_r)$ with transformations on the 2-sets of the following kind:

$$(i, k), (j, l) \mapsto (i, j), (k, l)$$

for any $(j, k | i, l) \in S$,

$$(i, k), (j, l) \mapsto (i, j), (k, l)$$

for any $(i, k | j, l) \in S$ such that $(i, k | j, l)$ is the only split of $\{i, j, k, l\}$ in S ,

$$(a, b) \mapsto (a, c), (c, b)$$

for any $a, b, c \in [n]$.

The proof is very similar to the one of Theorem 13; the only difference is that in the system (12) we have to consider also the inequalities induced (by replacing the D_{I_i} with the f_{I_i}) by the inequalities

$$D_{a,c} + D_{c,b} - D_{a,b} + \varepsilon > 0$$

for any distinct $a, b, c \in [n]$ and the inequalities

$$D_{a,b} + D_{c,d} < D_{a,c} + D_{b,d}$$

for any quartet $\{a, b, c, d\}$ in $[n]$ such that there is only one of its splits, $(a, b | c, d)$, in S .

References

- [1] A. Baldisserrri *Buneman's theorem for trees with exactly n vertices*, arXiv:1407.0048
- [2] A. Baldisserrri, E. Rubel *On graphlike k -dissimilarity vectors*, Ann. Comb., 18 (3) 356-381 (2014)
- [3] A. Baldisserrri, E. Rubel *Treelike families of multiweights*, arXiv:1404.6799

- [4] H-J Bandelt, M.A. Steel *Symmetric matrices representable by weighted trees over a cancellative abelian monoid*. SIAM J. Discrete Math. 8 (1995), no. 4, 517–525
- [5] C. Bocci, F. Cools *A tropical interpretation of m -dissimilarity maps* Appl. Math. Comput. 212 (2009), no. 2, 349–356
- [6] P. Buneman *A note on the metric properties of trees*, Journal of Combinatorial Theory Ser. B 17 (1974), 48–50
- [7] W. Carver *Systems of linear inequalities*, Ann. Math. 23 (3) 212–220 (1922)
- [8] H. Colonius, H.H. Schultze *Tree structure from proximity data*, British Journal of Mathematical and Statistical Psychology 34 (1981) 167–180
- [9] A. Dress, K. T. Huber, J. Koolen, V. Moulton, A. Spillner, *Basic phylogenetic combinatorics*. Cambridge University Press, Cambridge, 2012
- [10] S.Herrmann, K.Huber, V.Moulton, A.Spillner, *Recognizing treelike k -dissimilarities*, J. Classification 29 (2012), no. 3, 321–340
- [11] S.L. Hakimi, A.N. Patrinos *The distance matrix of a graph and its tree realization*, Quart. Appl. Math. 30 (1972/73), 255–269
- [12] S.L. Hakimi, S.S. Yau, *Distance matrix of a graph and its realizability*, Quart. Appl. Math. 22 (1965), 305–317
- [13] B. Iriarte Giraldo *Dissimilarity vectors of trees are contained in the tropical Grassmannian*, Electron. J. Combin. 17 (2010), no. 1
- [14] D. Levy, R. Yoshida, L. Pachter *Beyond pairwise distances: neighbor joining with phylogenetic diversity estimates*, Mol. Biol. Evol., 23 (2006), 491–498.
- [15] M. Nei, N.Saitou *The neighbor joining method: a new method for reconstructing phylogenetic trees* Mol. Biol. Evol. 4 (1987) no. 4, 406–425
- [16] L. Pachter, D. Speyer *Reconstructing trees from subtree weights*, Appl. Math. Lett. 17 (2004), no. 6, 615–621
- [17] E. Rubei *Sets of double and triple weights of trees*, Ann. Comb. 15 (2011), no. 4, 723–734
- [18] E. Rubei *On dissimilarity vectors of general weighted trees*, Discrete Math. 312 (2012), no. 19, 2872–2880
- [19] C. Semple, M. Steel, *Phylogenetics*. Oxford University Press, Oxford, 2003
- [20] J.M.S. Simoes Pereira *A Note on the Tree Realizability of a distance matrix*, J. Combinatorial Theory 6 (1969), 303–310
- [21] J.A. Studier, K.J. Keppler *A note on the neighbor-joining algorithm of Saitou and Nei* Mol. Biol. Evol. 5 (1988) no.6 729–731
- [22] K.A. Zaretskii *Constructing trees from the set of distances between pendant vertices*, Uspehi Matematicheskikh Nauk. 20 (1965), 90–92

Address: Dipartimento di Matematica e Informatica “U. Dini”, viale Morgagni 67/A, 50134 Firenze, Italia
E-mail address: rubei@math.unifi.it